UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/058,730	01/30/2002	Antonio Colmenarez	US010668 (020024) (3493-0	5696

24737 7590 03/07/2005

PHILIPS INTELLECTUAL PROPERTY & STANDARDS
P.O. BOX 3001
BRIARCLIFF MANOR, NY 10510

EXAMINER	
ALBERTALLI, BRIAN LOUIS	
ART UNIT	**PAPER NUMBER**
2655	

DATE MAILED: 03/07/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/058,730	COLMENAREZ ET AL.	
	Examiner	Art Unit	
	Brian L Albertalli	2655	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *30 January 2002*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-27* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-27* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *30 January 2002* is/are: a)☐ accepted or b)☒ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date *1/30/02, 6/4/03*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

Specification

1. The title of the invention is not descriptive. A new title is required that is clearly indicative of the invention to which the claims are directed.

The following title is suggested: --Speech Activity Detection Using Acoustic and Facial Characteristics in an Automatic Speech Recognition System--.

Drawings

2. New corrected drawings in compliance with 37 CFR 1.121(d) are required in this application because the drawings are informal and difficult to read. The corrected drawings are required in reply to the Office action to avoid abandonment of the application. The requirement for corrected drawings will not be held in abeyance.

Claim Rejections - 35 USC § 102

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

4. Claims 1-4, 6-10, 12-13, and 18-24 are rejected under 35 U.S.C. 102(e) as being anticipated by Basu et al. (U.S. Patent 6,594,629).

In regard to claim 1, Basu et al. disclose a speech recognition system (Fig. 1) comprising:

an acoustic detector for detecting speech utterances of a speaker (event detection module 28 performing audio event detection, column 15, line 66 to column 16, line 4);

a visual detector for detecting at least one facial characteristic associated with speech utterances of the speaker (event detection module performing mouth opening detection, column 15, lines 41-42);

a processing arrangement connected to be responsive to the acoustic and visual detectors for deriving a signal having first and second values respectively indicative of the speaker making and not making speech utterances such that the first value is derived in response to the acoustic detector detecting a finite, nonzero acoustic response while the visual detector detects at least one facial characteristic associated with speech utterances of the speaker (information from the audio event detection and the facial feature detection are used by event detection module 28 to turn search module 34 on or off, column 15, lines 19-22 and lines 59-65, column 16, lines 53-56; the event detection module 28 inherently only indicates the user is making speech if both the audio and visual detectors detect speech events); and

a speech recognizer for deriving an output indicative of the speech utterances as detected by the acoustic detector, the speech recognizer being connected to be responsive to the acoustic detector while the signal has the first value (search module

34 is turned on when a speech event is detected by event detection module 28, column 15, lines 59-61).

In regard to claim 2, Basu et al. disclose search module 34 is turned on when a speech event is detected (column 15, lines 59-61). The event detection is a combination of the detection of a visual speech event (mouth opening) as well as the detection of an audio event (column 16, lines 53-56). The signal output by search module 34, therefore, is the second value (off) in response to any of:

a) the acoustic detector not detecting a finite, nonzero acoustic response while the visual detector does not detect speech utterances of the speaker,

(b) the acoustic detector detecting a finite, nonzero acoustic response while the visual detector does not detect speech utterances of the speaker, and

(c) the acoustic detector not detecting a finite, nonzero acoustic response while the visual detector detects speech utterances of the speaker.

In regard to claims 3 and 6, Basu et al. disclose the processing arrangement includes a delay element (buffer) for assuring that the beginning of each speech utterance is coupled to the speech recognizer (speech is collected in the buffer, so that the speech can be sent for recognition if it is determined to be speech, column 15, lines 41-42 and lines 50-53).

In regard to claims 4 and 7, Basu et al. discloses the delay arrangement includes a memory element connected to be responsive to the acoustic detector (the buffer collects speech data), the memory element including a plurality of stages for storing sequential segments of the output of the acoustic detector, the delay arrangement being such that the contents of the memory element stage storing the beginning of a speech utterance are initially coupled to the speech recognizer (speech is collected in the buffer, so that the speech can be sent for recognition if it is determined to be speech, column 15, lines 41-42 and lines 50-53).

A buffer inherently collects sequential elements in a plurality of stages to ensure that the beginning of the elements are initially passed to the next stage. The buffer disclosed by Basu et al., therefore, inherently includes a plurality of stages for storing sequential segments of the output of the acoustic detector, the delay arrangement being such that the contents of the memory element stage storing the beginning of a speech utterance are initially coupled to the speech recognizer.

In regard to claims 8 and 9, Basu et al. disclose the delay arrangement (buffer) is arranged for assuring that upon the completion of each speech utterance the acoustic detector is decoupled from the speech recognizer (recognition is performed for each piece of buffered data, until no speech is uttered, column 15, lines 53-55).

In regard to claim 10, Basu et al. discloses the delay arrangement (buffer) includes a memory element connected to be responsive to the acoustic detector, the

memory element including a plurality of stages for storing sequential segments of the output of the acoustic detector, the delay arrangement being such that the contents of the memory element stage storing acoustic energy associated with the acoustic detector and which occurs upon completion of each speech utterance is prevented from being coupled to the speech recognizer (recognition is performed for each piece of buffered data, until no speech is uttered, column 15, lines 53-55).

A buffer inherently collects sequential elements in a plurality of stages to ensure that the beginning of the elements are initially passed to the next stage. The buffer disclosed by Basu et al., therefore, inherently includes a plurality of stages for storing sequential segments of the output of the acoustic detector, the delay arrangement being such that the contents of the memory element stage storing acoustic energy associated with the acoustic detector and which occurs upon completion of each speech utterance is prevented from being coupled to the speech recognizer.

In regard to claims 12 and 13, Basu et al. disclose the processing arrangement includes a face recognizer arranged for enabling the signal to have the first value in response to the speaker being at a predetermined orientation relative to the visual detector connected to be responsive to the visual detector (a face is identified as 'frontal' facing by frontal pose detector 20, column 7, lines 32-34 and column 15, lines 37-38).

In regard to claim 18, Basu et al. disclose a method of recognizing speech utterances of a speaker with an automatic speech recognizer responsive to acoustic speech utterances of the speaker comprising:

detecting acoustic energy having a spectrum associated with speech utterances (event detection module 28 performing audio event detection, column 15, line 66 to column 16, line 4),

detecting at least one facial characteristic associated with speech utterances of the speaker (event detection module performing mouth opening detection, column 15, lines 41-42), and

activating the automatic speech recognizer in response to the detected acoustic energy having a spectrum associated with speech utterances while the at least one facial characteristic associated with speech utterances of the speaker is occurring (information from the audio event detection and the facial feature detection are used by event detection module 28 to turn search module 34 on or off, column 15, lines 19-22 and lines 59-65, column 16, lines 53-56; the event detection module 28 inherently only indicates the user is making speech if both the audio and visual detectors detect speech events).

In regard to claim 19, Basu et al. disclose search module 34 is turned on when a speech event is detected (column 15, lines 59-61). The event detection is a combination of the detection of a visual speech event (mouth opening) as well as the detection of an audio event (column 16, lines 53-56). The method disclosed by Basu et

al. therefore comprises preventing activation of the automatic speech recognizer in response to any of:

(a) no acoustic energy having a spectrum associated with speech utterances being detected while no facial characteristic associated with speech utterances of the speaker is detected,

(b) acoustic energy having a spectrum associated with speech utterances being detected while no facial characteristic associated with speech utterances of the speaker is detected, and

(c) no acoustic energy having a spectrum associated with speech utterances being detected while at least one facial characteristic associated with speech utterances of the speaker is detected

In regard to claim 20, Basu et al. disclose assuring that the beginning of each speech utterance is coupled to the speech recognizer (with the buffer, column 15, lines 41-42 and lines 50-53).

In regard to claim 21, Basu et al. disclose the beginning of each speech utterance is assuredly coupled to the speech recognizer by:

(a) delaying the speech utterance (in the buffer, column 15, lines 41-42),

(b) recognizing the beginning of each speech utterance (with event detection module 28, using facial features and audio features, column 15, lines 41-42, column 15, line 66 to column 16, line 4, and column 16, lines 53-56), and

(c) responding to the recognized beginning of each speech utterance to couple the delayed speech utterance associated with the beginning of each speech utterance to the speech recognizer and thereafter sequentially coupling the remaining delayed speech utterances to the speech recognizer (column 15, lines 50-53).

In regard to claim 22, Basu et al. disclose assuring that no detected acoustic energy is coupled to the speech recognizer upon the completion of an utterance (recognition is performed for each piece of buffered data, until no speech is uttered, column 15, lines 53-55).

In regard to claim 23, Basu et al. disclose assurance that no detected acoustic energy is coupled to the speech recognizer upon the completion of a speech utterance is provided by:

(a) delaying the acoustic energy associated with the speech utterance (in the buffer, column 15, lines 41-42),

(b) recognizing the completion of each speech utterance (no more speech event, column 15, lines 53-55 and 59-61), and

(c) responding to the recognized completion of each speech utterance to decouple delayed acoustic energy occurring after the completion of each speech utterance from the speech recognizer (the process is only repeated until no more speech events are detected, then search module 34 is turned off, column 15, lines 53-55 and 59-61).

In regard to claim 24, Basu et al. disclose the at least one facial characteristic indicates the face of the speaker has a predetermined orientation relative to a detector involved in the step of detecting the at least one facial characteristic (a face is identified as 'frontal' facing, column 15, lines 37-38).

Claim Rejections - 35 USC § 103

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

6. Claims 5 and 11 are rejected under 35 U.S.C. 103(a) as being unpatentable over Basu et al.

Basu et al. does not disclose that the buffer is a ring buffer.

Official notice is taken that it is notoriously well known and recognized in the art to use ring buffers when buffering data because they are easy to implement and ensure the buffer stays at a fixed size.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Basu et al. so that the buffer was a ring buffer because ring buffers are easy to implement and ensure the buffer stays at a fixed size.

7. Claims 14-17 and 25-27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Basu et al. (U.S. Patent 6,594,629, hereinafter Basu 1), in view of Basu et al. (U.S. Patent 6,219,640, hereinafter Basu 2).

In regard to claim 14, Basu 1 discloses the system has applications in speaker detection in an audience, as well as speaker recognition (identifying who is speaking), and refers to the Basu 2 reference (column 17, lines 11-15).

Basu 1 does not explicitly disclose distinguishing the faces of a plurality of speakers and enabling the speaker to have the first value in response to a speaker having a recognized face.

Basu 2 discloses a system for identifying a speaker that includes a face recognizer (Fig. 1, face recognition 24) that distinguishes from a plurality of speakers (identifies the person speaking, column 6, lines 46-50).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Basu 1 to identify who was speaking using a face recognizer and enable the speech recognizer when in response to a speaker having a recognized face, in order to get accurate recognition results when the user was in a crowd.

In regard to claim 15, Basu 1 discloses the system has applications in speaker detection in an audience, as well as speaker recognition (identifying who is speaking), and refers to the Basu 2 reference (column 17, lines 11-15).

Basu 1 does not explicitly disclose including a speaker identity recognizer to be responsive to the acoustic detector, to distinguish speech patterns of a plurality of

speakers and enable the signal to have the first value in response to the speaker having a recognized speech pattern.

Basu 2, discloses a speaker identity recognizer (speaker recognizer 16) arranged for distinguishing speech patterns of a plurality of speakers (column 5, lines 29-34).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Basu 1 to identify who was speaking using a speaker recognizer enable the speech recognizer in response to the speaker's voice being recognized, in order to provide a second means for confirming the identity of the speaker, provide a backup if the facial recognizer had difficulty identifying the user.

In regard to claims 16 and 17, Basu 1 discloses the system has applications in speaker detection in an audience, as well as speaker recognition (identifying who is speaking), and refers to the Basu 2 reference (column 17, lines 11-15).

Basu 1 does not disclose the processing arrangement is arranged for causing the signal to have the first value in response to the speaker having a recognized face matched with a recognized speech pattern of the same speaker.

Basu 2 discloses a speaker is identified when a face is matched with a recognized speech pattern of the same speaker (joint identification module 30, column 8, lines 43-47).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Basu 1 to enable the speech recognizer when the recognized face

matched the recognized speech pattern of the same speaker, in order to improve the speaker recognition accuracy, as taught by Basu 2 (column 3, lines 32-36).

In regard to claims 25 and 26, discloses the method has applications in speaker detection in an audience, as well as speaker recognition (identifying who is speaking), and refers to the Basu 2 reference (column 17, lines 11-15).

Basu 1 does not explicitly disclose distinguishing the faces of a plurality of speakers and enabling the speaker to have the first value in response to a speaker having a recognized face.

Basu 2 discloses a method for identifying a speaker that includes a face recognizer (Fig. 1, face recognition 24) that distinguishes from a plurality of speakers (identifies the person speaking, column 6, lines 46-50).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Basu 1 to identify who was speaking using a face recognizer and enable the speech recognizer when in response to a speaker having a recognized face, in order to get accurate recognition results when the user was in a crowd.

In regard to claim 27, Basu 1 discloses the method has applications in speaker detection in an audience, as well as speaker recognition (identifying who is speaking), and refers to the Basu 2 reference (column 17, lines 11-15).

Basu 1 does not disclose storing images of the faces and speech patterns during a training period and comparing the stored images and speech patterns with images of the face of the speaker and the speech pattern of the speaker.

Basu 2 discloses storing:

(1) images of the faces of a plurality of speakers (column 7, lines 27-28), and

(2) the speech patterns of the same plurality of speakers during at least one training period (column 6, lines 17-20, see also column 14, lines 10-12); and

performing the distinguishing steps by comparing the stored images and speech patterns with images of the face of the speaker and the speech pattern of the speaker (joint identification, column 8, lines 43-47)

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Basu 1 to store images of the faces of a plurality of speakers and the speech patterns of the same plurality of speakers during a training period, since these are necessary to later identify the users. Furthermore, it also would have been obvious to one of ordinary skill in the art at the time of invention to enable the speech recognizer when the recognized face matched the recognized speech pattern of the same speaker, in order to improve the speaker recognition accuracy, as taught by Basu 2 (column 3, lines 32-36).

Conclusion

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. De Cuetos et al. (*Audio-Visual Intent-to-Speak Detection for*

Human-Computer Interaction and U.S. Patent 6,754,373) disclose a system that turns on a microphone for a speech recognizer when a frontal facial image is detected. Fiex et al. (U.S. Patent 4,449,189), Brunelli et al. (U.S. Patent 5,412,378), Bakis et al. (U.S. Patent 6,219,639) disclose systems that use visual and speech cues to identify a user. Strubbe et al. (U.S. Patent 6,721,706) disclose a chatterbot that varies its responses according to the combined input from a camera and speech recognizer.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Brian L Albertalli whose telephone number is (703) 305-1817. The examiner can normally be reached on Mon - Fri, 8:00 AM - 5:30 PM, every second Fri off.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Talivaldis Smits can be reached on (703) 305-3011. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

BLA 3/1/05

TĀLIVALDIS IVARS ŠMITS
PRIMARY EXAMINER